

CONDITIONAL BELIEF STRUCTURES

JÜRIG KOHLAS

*Institute for Automation and Operations Research
University of Fribourg
Switzerland*

1. INTRODUCTION

The mathematical theory of evidence (Shafer et al. [9]) has recently found much interest as an approach to treat uncertainty in expert and knowledge-based systems. Although the theory is very promising, there are not yet many practical applications. Modeling practice has still to be developed. This is a crucial task in view of facilitating the application of evidential modeling. It is the aim of this paper to discuss an important element of evidential modeling—conditional belief—within the scope of the mathematical theory of evidence.

In a given problem domain, there exist in general *domain dependent, structural knowledge* about various relations between different aspects or elements of the problem. On the other hand, there are several sources of information or knowledge, which produce evidence about various facets of the problem in a given context or situation. This observational evidence must then be combined with the structural knowledge.

Any model within the mathematical theory of evidence is based on a *frame of discernment* Θ , a finite set whose elements represent the possible outcomes or results. Often the reasoning is about different objects, different attributes or properties of an object, or more generally about different variables related to a problem. It is then natural to consider Θ as a product set formed by several factor frames Θ_i , each one representing the possible values of one variable. The structural knowledge in such models is then contained in relations between different variables or different factor frames Θ_i . Such relations may take the form of *conditional belief structures*. Intuitively, these structures describe the beliefs about propositions in Θ_j , given certain propositions in Θ_i . The exact mathematical definition of these conditional belief structures is the first goal of this paper.

Observational evidence coming from a knowledge source pertains often to one of the factor frames Θ_i and can be modeled by a belief function on Θ_i . The belief functions representing observational evidences have then to be combined with the conditional belief structures representing structural knowledge. The analysis of this combination of observational and structural knowledge is the second goal of this paper.

Lowrance and Garvey [5], Lowrance [4], and Garvey [2] describe applications which are modeled along the above lines. Their conditional belief structures reduce, however, to simple and fixed *compatibility relations* (compare Section 3 below). The concepts introduced in this paper generalize their approach. Lu and Stephanou [6] discuss another approach which is generalized in this paper.

The outline of this paper is as follows: In Section 2, a short introduction to the mathematical theory of evidence is given. In Section 3, two simple, but interesting examples, are discussed as an introduction into the subject of the paper. In Section 4, conditional belief structures are introduced on the basis of simple compatibility relations. Some applications are given, notably a discussion of *inference networks*. In Section 5, more general relations are considered and the general definition of conditional belief structures is introduced. The propagation of belief from one frame Θ_i to another frame Θ_j by conditional belief structures is discussed in detail. As an application, the case of independent “if . . . then . . .” rules, i.e., the model of Lu and Stephanou [6] is analyzed.

2. ELEMENTS OF THE MATHEMATICAL THEORY OF EVIDENCE

The central notion of Shafer’s mathematical theory of evidence is the *belief function* (Shafer [8]). It expresses mathematically the belief induced by a given body of evidence on possible outcomes or results and propositions formed by them. Let Θ be a *finite* set, whose elements represent the possible outcomes. Θ is called the *frame of discernment*. The evidence available concerning the possible outcomes is regarded as a message, whose interpretation is uncertain. Suppose there are $i = 1, 2, \dots, m$ different possible interpretations and that the i th interpretation has probability p_i of being the correct one. These probabilities are positive and sum up to one:

$$\sum_{i=1}^m p_i = 1, \quad p_i > 0, \quad (1)$$

Now, suppose the i th interpretation of the available evidence says that the true (and unknown) outcome lies within the subset $A_i \subset \Theta$. There is thus a probability of p_i that A_i is the right interpretation of the evidence.

Now fix any subset $A \subset \Theta$. It can be concluded that the true outcome lies

within A at least, if any one of the interpretations A_i of the available evidence with $A_i \subset A$ is the correct one. The probability of this event can be considered as a *degree of belief* induced into proposition (or event) A by the available evidence on the possible outcomes. We define thus

$$Bel(A) = \sum_{A_i \subset A} p_i, \quad (2)$$

where Bel denotes the degree of belief. Note that Bel is an application from 2^Θ to $[0,1]$, but that it is—in general—not additive, i.e., $Bel(A \cup B) \neq Bel(A) + Bel(B)$, even if A and B are disjoint. The sets A_i are called *focal sets* and their probabilities p_i *basic probability commitments*. It is convenient to introduce a function $m(B) = p_i$ if $B = A_i$ and $m(B) = 0$ otherwise. Then Eq. (2) can also be written as

$$Bel(A) = \sum_{B \subset A} m(B). \quad (3)$$

The introduction of belief functions as given here using the model of different possible interpretations of an evidence was proposed by Shafer and Tversky [10]. A slightly different one goes back to Dempster [1]; see also Shafer [8].

The more belief that is assigned to the complement A^c of a proposition A , the more doubt is cast upon A itself. Therefore,

$$Dou(A) = Bel(A^c) \quad (4)$$

is called the *degree of doubt* of A . The less doubt there is in a proposition A , the more plausible appears this proposition. That is

$$Pl(A) = 1 - Dou(A) = 1 - Bel(A^c) \quad (5)$$

is called the *degree of plausibility* in A . It is easy to see that always

$$Bel(A) \leq Pl(A). \quad (6)$$

Bel and Pl are therefore sometimes considered as the upper and lower probability limits within which the true (but unknown) probability must be contained (e.g., Dempster [1]).

Often an evidence is to be interpreted as either pointing to a focal set $A \subset \Theta$ (probability p) or else containing no information whatsoever (probability $1 - p$). In the latter case, the corresponding focal set is the whole frame of discernment Θ . The corresponding belief function is then called a *simple* belief function. Belief functions whose focal sets are nested, $A_{i-1} \subset A_i$, for $i = 2, \dots, m$, are called *consonant*.

There might be different evidences stemming from different sources, but

pertaining to the same frame of discernment. Suppose there are two bodies of evidence, $j = 1, 2$, each one inducing a belief function Bel_j with focal sets A_{ij} and basic probability commitments $p_{ij}, i = 1, 2, \dots, m_j, j = 1, 2$. How can these two bodies be fused into a single, combined body of evidence, inducing a combined belief function? Consider interpretation i of the first body and interpretation k of the second body. If both are the correct interpretations of the two corresponding bodies of evidence, then their combined interpretation must lead to the focal set $A_{i1} \cap A_{k2}$. If, furthermore, the two bodies and their possible interpretations are considered to be independent in the usual stochastic sense, then the above combined interpretation has the probability $p_{i1}p_{k2}$ to be correct.

It cannot be excluded that the intersection $A_{i1} \cap A_{k2}$ is *empty*. This is a contradiction in the sense that not both interpretations can be correct at the same time. In other words, their combination is no possible combination of interpretations and must be excluded. This means that the probabilities of combinations must be conditioned on the event that these combinations are possible. Furthermore, different interpretations i and k may lead to the same intersection $A_{i1} \cap A_{k2}$. Thus, the combined belief function is defined by the following basic probability commitments:

$$m(B) = \sum_{A_{i1} \cap A_{k2} = B} m(A_{i1})m(A_{k2})/k, \quad (7)$$

where

$$k = 1 - \sum_{A_{i1} \cap A_{k2} = \emptyset} m(A_{i1})m(A_{k2}). \quad (8)$$

This is *Dempster's rule* for the combination of *independent* bodies of evidence (Dempster [1]). The new, combined belief function is denoted by $Bel_1 \oplus Bel_2$. It is easy to verify that this operation of combination is *commutative* and *associative*. As mentioned in the introduction, the frame of discernment is often a product space. Suppose $\Theta = \Theta' \times \Theta''$. In many cases, a belief function is defined on one of the factors, say Θ' . It can then be extended to Θ by defining the focal sets $A_i \times \Theta''$. This is called the *vacuous extension*. Sometimes one wants to restrict a belief function defined on Θ to one of the factors, say Θ'' . This is accomplished by projecting the focal sets of the belief function to the factor and taking these projections as new focal sets (it is possible that several focal sets project to the same set, then their probabilities have to be summed up). These two operations are special cases of the corresponding operations related to the refining and coarsening of frames as discussed by Shafer [8]. For that and other aspects of the mathematical theory of evidence, we refer to Shafer's book [8].

3. TWO SIMPLE EXAMPLES

3.1. Fingerprints

Both examples which are analyzed in this section are taken from Pearl [7], who treats them with the classical Bayesian analysis, whereas they are discussed here within the framework of evidential modeling.

Assume in the first example that in a certain trial there are some suspects, one of whom might have committed a murder and that the murder weapon, showing some fingerprints, has been found by the police. It is believed that the killer was the last to hold the weapon. In this example, there are two variables or frames of discernment Θ' and Θ'' related to the two questions "who was the murderer?" and "who last held the weapon?". Let $\theta_1, \theta_2, \dots, \theta_n$ denote the suspects, but assume that it is not excluded that somebody else committed the murder, a possibility denoted by θ' . Then $\Theta' = \{\theta_1, \dots, \theta_n, \theta'\}$. In the same way, Θ'' contains also the suspects θ_i , but again assume that it is not excluded that somebody else last held the weapon, a possibility denoted this time by θ'' , hence $\Theta'' = \{\theta_1, \dots, \theta_n, \theta''\}$. If it is in fact true that the murderer was the last to hold the weapon, then only the points $C_1 = \{(\theta_1, \theta_1), (\theta_2, \theta_2), \dots, (\theta_n, \theta_n), (\theta', \theta'')\} \subset \Theta' \times \Theta''$ are possible. Such a subset C_1 of the product set $\Theta' \times \Theta''$ is called a *compatibility relation*. It represents a *restriction* on the combined possibilities of both frames. But it is not sure that it was the killer who last held the weapon. Then all points of $C_2 = \Theta' \times \Theta''$ are possible. Suppose a probability of p can be assigned to the possibility that the murderer was in fact the last to hold the weapon. Then $m(C_1) = p$, $m(C_2) = 1 - p$ defines a simple belief function Bel on $\Theta' \times \Theta''$ which represents the belief that the killer held the weapon last.

This construction can also be interpreted in the following way: Given that θ_i is in fact the murderer, then there is a simple belief function on Θ'' , focusing a probability p to $\{\theta_i\}$, that the fingerprints of θ_i are on the weapon. Or symmetrically: given that θ_i was the last to hold the weapon, then this induces a simple belief function on Θ' , focusing the probability p on $\{\theta_i\}$, that θ_i is in fact the murderer. That is why Bel may be called a *conditional belief structure* describing a relation between the two frames Θ' and Θ'' .

The laboratory which analyses the fingerprints is a knowledge source which produces evidence about who last held the weapon. The results of the analysis may be represented by a belief function Bel'' on Θ'' . Using the vacuous extension of Bel'' to $\Theta' \times \Theta''$, $Bel \oplus Bel''$ can be computed by Dempster's rule, and the restriction Bel' of $Bel \oplus Bel''$ to Θ' represents the resulting belief about who could have been the killer.

Suppose (i) that the laboratory produces a simple belief function Bel'' focusing a probability q on $\{\theta_i\}$ for some i . Then it can be easily verified that this charges θ_i with a belief of $Bel'(\{\theta_i\}) = pq$ to be the murderer. Suppose (ii) that the laboratory produces a simple belief focusing a probability q on $\{\theta''\}$. Then it follows that the suspects $\theta_i, i = 1, 2, \dots, n$, are discharged from the

accusation by a small plausibility $Pl'(\{\theta_1, \dots, \theta_n\}) = 1 - pq$ for any one of them being the killer. Other cases may of course be invented and treated and other knowledge sources yielding evidence in favor or against the suspects in Θ' (such as alibis) may be introduced and combined with the evidence from the analysis of fingerprints on the weapon.

3.2. Alarm System

Pearl [7] considers the following small story: “Mr. Holmes received a phone call from his neighbor notifying him that she heard a burglar alarm sound from the direction of his home. As he is preparing to rush home, Mr. Holmes recalls that recently the alarm had been triggered by an earthquake. Driving home, he hears a radio newscast reporting an earthquake 200 miles away.” The question is, should Mr. Holmes be worried about a burglary?

In this case, three variables or factor frames must be considered: $B = \{b, \text{not } b\}$ concerning the question whether there is or is not a burglary, $E = \{e, \text{not } e\}$ regarding the question whether there is or is not an earthquake in the region, and $A = \{a, \text{not } a\}$ representing the possibility of an alarm. Here causal relationships of the form “if b then probably (hopefully) a ”, “if e then possibly a ”, “if neither b nor e then surely not a ” have to be modeled. This means that the two variables B and E together causally influence A . Four possible cases have to be distinguished: (1) Both burglary and earthquake trigger alarm. This restricts the possible cases in the product set $B \times E \times A$ to $C_1 = \{(b, e, a), (b, \text{not } e, a), (\text{not } b, e, a), (\text{not } b, \text{not } e, \text{not } a)\}$. (2) Burglary triggers alarm, but earthquake does not. Similarly, (3) burglary does not trigger alarm, but earthquake does. In these two cases, the following restrictions hold: $C_2 = \{(b, e, a), (b, \text{not } e, a), (\text{not } b, e, \text{not } a), (\text{not } b, \text{not } e, \text{not } a)\}$; $C_3 = \{(b, e, a), (b, \text{not } e, \text{not } a), (\text{not } b, e, a), (\text{not } b, \text{not } e, \text{not } a)\}$. Finally (4), it is possible that neither burglary nor earthquake triggers an alarm, but it is not excluded that alarm is caused by another source. In this case, all combinations in $C_4 = B \times E \times A$ are possible. C_1 to C_4 are compatibility relations between B , E , and A . Let p and q be the probabilities that a burglary triggers an alarm and that an earthquake triggers an alarm, respectively, and suppose that these two events are stochastically independent. Then the four cases have the four probabilities $pq, p(1 - q), (1 - p)q$, and $(1 - p)(1 - q)$. These are the probabilities committed to the four sets C_1 to C_4 in $B \times E \times A$. This defines a belief function Bel' on the product set $B \times E \times A$.

This belief function can again be considered as a *conditional belief structure*. In fact, given that there is at the same time a burglary and an earthquake (b, e) , then by C_1 , C_2 , and C_3 an alarm will be triggered with probability $pq + p(1 - q) + (1 - p)q = 1 - (1 - p)(1 - q)$, whereas according to C_4 it is not sure that an alarm is triggered (probability $(1 - p)(1 - q)$). This represents a belief function on A . In the same way, conditional beliefs of an alarm given $(b, \text{not } e)$ and $(\text{not } b, e)$ are defined. On the other hand, given that there

is an alarm, there is by C_1 a burglary or an earthquake (probability pq), by C_2 there is a burglary (probability $p(1 - q)$), and by C_3 an earthquake (probability $(1 - p)q$). By C_4 , it is also possible that neither an earthquake nor a burglary caused the alarm. Thus, the condition $a \in A$ induces also a conditional belief function on $B \times E$.

Consider now the phone call from the neighbor reporting an alarm. This can be represented as a simple belief function focusing a belief 1 to $a \in A$. Combining this simple belief function with Bel' by Dempster's rule gives a belief function on $B \times E \times A$ with focal sets $\{(b, e, a), (b, \text{not } e, a), (\text{not } b, e, a)\}; \{(b, e, a), (b, \text{not } e, a)\}; \{(b, e, a), (\text{not } b, e, a)\}$, and $B \times E \times \{a\}$ and basic probability numbers $pq, p(1 - q), (1 - p)q$, and $(1 - p)(1 - q)$, respectively. This defines the following beliefs: $Bel(\text{burglary or earthquake}) = 1 - (1 - p)(1 - q)$, $Bel(\text{burglary}) = p(1 - q)$.

q may be nearly 1, i.e., an alarm system may be very sensitive to earthquakes. Then there is at first sight a surprisingly small belief in a burglary. Suppose, e.g., $p = 0.9, q = 0.8$, then $Bel(\text{burglary}) = 0.18$ only. That is because there is another evidence to be combined: if there is no radio message about the occurrence of an earthquake, then there is strong evidence, based on past experience on the rarity of earthquakes, that there is in fact no earthquake. This can be represented by a simple belief function focusing s (near to 1) to $\text{not } e \in E$. If this supplementary evidence is combined by Dempster's rule with the above belief function, then there results a belief into burglary to $Bel(\text{burglary}) = (pqs + p(1 - q))/(1 - (1 - p)qs)$. This gives, for $s = 0.99$, a belief of 0.96959 as compared to 0.18 above. This result shows that it is important to include always all the available evidence.

But assume now that there is a radio message about an earthquake. Then this is strong evidence in favor of $e \in E$, represented by a belief function focusing r (near to 1) to $e \in E$. If this is combined by Dempster's rule with the above belief function, then the belief in burglary remains the same, $Bel(\text{burglary}) = p(1 - q) = 0.18$, independently of r . But it is of course drastically reduced compared to the case where there is no evidence about an earthquake.

4. SIMPLE COMPATIBILITY RELATIONS

4.1. Two Variables

The examples of Section 3 show how to define conditional belief structures linking two or more variables together. In this subsection, the particular case of *two* variables will be considered. Consider, therefore, the two frames of discernment Θ' and Θ'' corresponding to the two variables. Each subset $C \subset \Theta' \times \Theta''$ of the product set is a *compatibility relation* between the two frames or variables which restricts the possible values which the two variables may take at the same time. For any $\theta' \in \Theta'$, define the set $C(\theta') = \{\theta'' \in \Theta'' : (\theta', \theta'') \in C\}$ of values in Θ'' compatible with θ' in Θ' . It will not be excluded that $C(\theta')$ may be

empty; θ' is then an impossible value. In the same way, $C(\theta'')$ is defined as the set of values $\theta' \in \Theta'$ compatible with θ'' .

A conditional belief structure on $\Theta' \times \Theta''$ is defined by a belief function Bel with compatibility relations $C_i \in \Theta' \times \Theta''$, $i = 1, 2, \dots, m$, as focal sets and basic probability commitments p_i to these sets. Given a $\theta' \in \Theta'$, there is then a conditional belief function on Θ'' with focal sets $C_i(\theta')$ and basic probabilities p_i (assuming no $C_i(\theta')$ is empty). Symmetrically, given a $\theta'' \in \Theta''$, there is a conditional belief function on Θ' with focal sets $C_i(\theta'')$ and basic probabilities p_i . This allows the consideration of uncertain rules of the form “if θ' then $C_i(\theta')$ ” with some probability p_i . But many times one wants to consider more generally propositions P' in Θ' and uncertain rules of the form “if P' then Q'' ” with some probability, where Q'' is a proposition in Θ'' . This case will be discussed in the next section.

If Bel' is a belief function on Θ' , then it may be extended to $\Theta' \times \Theta''$ by the vacuous extension, then combined with Bel using Dempster's rule and finally restricted in Θ'' , resulting in a belief function Bel'' on Θ'' representing the belief induced on Θ'' by Bel' and the conditional belief structure Bel . Of course, belief may also be propagated from Θ'' to Θ' in a similar way.

4.2. Many Variables

Consider now more generally the case of $n \geq 2$ variables or frames of discernment $\Theta_i, i = 1, 2, \dots, n$. Let $I_n = \{1, 2, \dots, n\}$ be the corresponding index set. If $W \subset I_n$ contains at least two elements, then a subset

$$C \subset \prod_{j \in W} \Theta_j \quad (9)$$

is a *compatibility relation* between variables or frames $j \in W$ and restricts possible values of tuples $(\theta_j, j \in W)$. A *conditional belief structure* between variables $j \in W$ is a belief function whose focal sets are compatibility relations C_i , $i = 1, 2, \dots, m$, between variables $j \in W$ with basic probability commitments p_i .

If H is a family of several subsets $W \subset I_n$ (all of them containing at least two elements), then for any $W \in H$ a conditional belief structure Bel_W can be considered. It will generally be assumed that the probabilities of the different conditional belief structures are stochastically independent. If, furthermore, belief functions on one or several frames Θ_i are given, then these beliefs may be combined with the conditional belief structures using Dempster's rule. Systems of this kind have been considered by Kong [3]. Shafer et al. [9] studied special cases of this structure and indicated computationally efficient methods to combine evidence. In the following subsection, another important special case is presented.

4.3. Inference Networks

Consider a set of propositions p_k , $k = 1, 2, \dots, n$, linked by rules “if p_i then p_j ” for some couples (i, j) . The propositions p_k form the nodes V and the rules “if p_i then p_j ” the *oriented* edges E of a graph $G = (V, E)$. G is called an *inference network*. It will be assumed that the rules are not fully reliable in the sense that a condition p_i implies the conclusion p_j only with some probability m_{ij} smaller than 1, whereas with probability $(1 - m_{ij})$ p_i implies neither p_j nor not p_j . To represent such a situation, consider frames $\Theta_i = \{p_i, \text{not } p_i\}$. If a rule “if p_i then p_j ” holds then only the couples $\{(p_i, p_j), (\text{not } p_i, p_j), (\text{not } p_i, \text{not } p_j)\} \subset \Theta_i \times \Theta_j$ are possible, otherwise all of $\Theta_i \times \Theta_j$ is possible. Let us denote the former set by $(\text{not } p_i) \text{ or } p_j$. Then the unreliability of the rule “if p_i then p_j ” can be described by the simple belief function Bel_{ij} on $\Theta_i \times \Theta_j$ focusing the probability m_{ij} to $(\text{not } p_i) \text{ or } p_j$. This corresponds to the general case discussed in the previous subsection where H is here equal to E and all conditional belief structures are between pairs of frames Θ_i and Θ_j .

Suppose furthermore that there are some independent, simple belief functions Bel_i for some propositions p_i , $i \in M \subset I_n$, which assign probabilities m_i to p_i . Let Bel denote the total combined belief of the simple beliefs about propositions p_i , $i \in M$ and the conditional belief structures associated to the rules of the network,

$$Bel = \left(\bigoplus_{i \in M} Bel_i \right) \oplus \left(\bigoplus_{(i,j) \in E} Bel_{ij} \right). \quad (10)$$

It is possible to interpret m_i as the probability that node p_i in G is intact or available and m_{ij} in the same sense, that edge (i, j) in G is intact or available. G is then considered as a network where some nodes and edges may fail. For any node q in G , there is a probability $p_{\text{link}}(q)$ that at least one node p_i , $i \in M$, together with a path from p_i to q in G is available. The following theorem establishes a relation between total belief Bel and the network reliability:

THEOREM 4.1: *For any node q of the graph G ,*

$$Bel(q) = p_{\text{link}}(q). \quad (11)$$

PROOF: The focal sets of Eq. (10) are obtained by taking any subsets W of M and U of E and intersecting p_i , $i \in W$, and $(\text{not } p_k) \text{ or } p_j$, $(k, j) \in U$. Suppose then first there exists a path from a $w \in W$ (we write $w \in W$ instead of $w = p_i$, $i \in W$) to q within the *subgraph* (V, U) of G ($U \subset E$). Consider

$$q_U = \bigcap_{(k,j) \in U} ((\text{not } p_k) \text{ or } p_j) \quad (12)$$

and let $w, p_1, p_2, \dots, p_r, q$ define a path from w to q within (V, U) . Then

$$\begin{aligned} q_U &\subset ((\text{not } w) \text{ or } p_1) \cap ((\text{not } p_1) \text{ or } p_2) \cap \dots \cap ((\text{not } p_r) \text{ or } q) \\ &\subset (\text{not } w) \text{ or } q \end{aligned} \quad (13)$$

and thus

$$q_{(W,U)} = \left(\bigcap_{i \in W} p_i \right) \cap q_U \subset w \cap ((\text{not } w) \text{ or } q) \subset q. \quad (14)$$

Hence, every focal set of Bel in Eq. (10) which defines a subgraph containing a path from some $w = p_i$, $i \in M$ to q implies q . This means that $Bel(q) \geq p_{\text{link}}(q)$.

Suppose now, on the other hand, that either $W = \emptyset$ or there is no path from a $w \in W$ to q within the subgraph (V, U) . If $W = \emptyset$, then $q_{(W,U)} = q_U$ may be developed into a disjunction of literals $x_1 x_2 \dots$, where each x_i is either a p_j or a $\text{not } p_j$. But obviously at least one literal of this disjunction does not contain q and therefore q_U does not imply q .

If W is not empty, but there is no path from a $w \in W$ to q within (V, U) , then consider the subset of edges U' of U which are on some path leading to q . U' may be empty. But if U' is not empty, then $q_{U'}$ may again be developed into a disjunction of literals. None of these literals contains a $w \in W$ because there is no path from w to q . At least one of these literals does not contain q .

Let $U'' = U - U'$. If U'' is not empty then consider once again the development of $q_{U''}$ into the disjunction of literals. None of these literals contain q and for any $w \in W$ at least one of the literals does not contain $(\text{not } w)$. Then $q_U = q_{U'} \cap q_{U''}$. Combining the two disjunctions of literals corresponding to $q_{U'}$ and $q_{U''}$ yields the disjunction of literals corresponding to q_U . There is then for any $w \in W$ a literal containing neither $(\text{not } w)$ nor q . Thus, q_U does not imply $(\text{not } w) \text{ or } q$ and this implies that $q_{(W,U)}$ does not imply q . Hence, $Bel(q) = p_{\text{link}}(q)$. ■

The computation of $Bel(q)$ is in general no easy task, except in some special cases. If G contains no cycles then the method described in Shafer et al. [9] can be applied.

5. GENERALIZED COMPATIBILITY RELATIONS

5.1. Modeling Rules

This section is primarily motivated by the desire to model uncertain rules like “if P then probably Q ” within the mathematical theory of evidence. More generally assume $P \subset \Theta'$ is some proposition discerned by the frame of discernment Θ' . If P is known to be true, then some (conditional) belief function Bel_P^θ on Θ'' may describe what follows from P . Even, more generally, if P_k , $k = 1, 2, \dots$, is a family of propositions in Θ' , then for any k there is a (conditional) belief function Bel_k on Θ'' describing the belief induced by P_k . If Bel represents a belief function on Θ' , then one wants, of course, to combine this belief with the conditional belief structure defined by Bel_k to propagate belief from Θ' to Θ'' . This is only a sketch of what one wants to achieve. Several attempts have been reported in the literature to model that within the framework of the

mathematical theory of evidence (Lu and Stephanou [6]), but none of them was entirely satisfactory. In this subsection, a rigorous, formal solution to this problem is proposed.

The simple compatibility relations and conditional belief structures discussed in the last section are by themselves not yet sufficient to model the kind of rules mentioned above. But they offer, nevertheless, the base for the solution. Let $P_k, k = 1, 2, \dots, n, P_k \subset \Theta'$, and $Q_j, j = 1, 2, \dots, m, Q_j \subset \Theta''$ be two families of arbitrary propositions in Θ' and Θ'' , respectively. Note that it is not assumed that the P_k (nor the Q_j) are mutually disjoint. Let I_n and I_m be the index sets of the two families. Consider a compatibility relation $R \subset I_n \times I_m$ between propositions of the two families. So if $(k, j) \in R$, then the two propositions P_k and Q_j are compatible; i.e., if P_k is known to be true, then Q_j is possible and vice versa. We call this a *generalized compatibility relation* which will be denoted by R .

Let $R(k) = \{j \in I_m: (k, j) \in R\}$ and $R(j) = \{k \in I_n: (k, j) \in R\}$. Then together with $\theta' \in P_k$, only points $\theta'' \in Q_j$ with $j \in R(k)$ are possible or symmetrically, to $\theta'' \in Q_j$, only points $\theta' \in P_k, k \in R(j)$, are compatible. Thus, a generalized compatibility relation between propositions in Θ' and Θ'' imposes constraints on the possible pairs (θ', θ'') which are described by the *simple* compatibility relation

$$C(R) = \bigcup_{(j,k) \in R} P_k \times Q_j \subset \Theta' \times \Theta''. \quad (15)$$

If C is an arbitrary, simple compatibility relation between Θ' and Θ'' , then the two family of propositions $\{\theta' \in \Theta'\}$ and $\{\theta'' \in \Theta''\}$ may be considered, and C then defines a generalized compatibility relation C' between these two families of propositions such that clearly $C(C') = C$. That's why one may speak of *generalized* compatibility relations. As a further example, note that a rule like "if P then Q " ($P \subset \Theta', Q \subset \Theta''$) is defined by the generalized compatibility relation $R = \{(P, Q), (\text{not } P, \Theta'')\}$ between the families of propositions $\{P, \text{not } P\}$ and $\{Q, \Theta''\}$. Associated with this generalized compatibility relation is the simple compatibility relation $C(R) = P \times Q \cup (\text{not } P \times \Theta'')$, which we will note in a convenient abbreviation as $(\text{not } P) \text{ or } Q$. Note furthermore that the generalized compatibility relation $\{(\text{not } P, \text{not } Q), (\Theta', Q)\}$ leads to the same simple compatibility relation $(\text{not } (\text{not } Q))$ or $(\text{not } P) = (\text{not } P) \text{ or } Q$. As the simple compatibility relation $C(R)$ is what really matters to treat generalized compatibility relations (see below), the two generalized compatibility relations above may be considered as *equivalent*.

Suppose a generalized compatibility relation R between propositions in Θ' and Θ'' is given and that a proposition $P \subset \Theta'$ is known to be true. This restricts the possible values in $\Theta' \times \Theta''$ to $(P \times \Theta'') \cap C(R)$. The projection of this set to Θ'' determines the set of possible points in Θ'' (i.e., points which are compatible with P). In the case of a rule "if P then Q " as it is modeled above, it

follows that a proposition P' which implies $P(P' \subset P)$ projects to Q , whereas any other proposition $(P' \cap \text{not } P \neq \emptyset)$ projects to Θ'' . This is exactly as it must be.

To introduce uncertainty into the rules, consider, in a general approach for $i = 1, 2, \dots, r$, different generalized compatibility relations R_i and suppose that R_i is selected with some given probability p_i (the p_i sum to 1). This then defines a belief function Bel on $\Theta' \times \Theta''$ with focal sets $C(R_i)$ and basic probability numbers p_i , hence a conditional belief structure on $\Theta' \times \Theta''$ in the sense of Subsection 3.1. This is how uncertain rules must be represented within the framework of the mathematical theory of evidence. If there is a belief function Bel' given on Θ' , then it may be extended to $\Theta' \times \Theta''$ by the vacuous extension and then be combined with Bel by Dempster's rule. The belief thereby propagated to Θ'' is obtained by the restriction of $Bel' \oplus Bel$ to Θ'' . In a symmetrical way, belief Bel'' on Θ'' may be transported to Θ' .

This will be illustrated by a few very simple examples. Consider first an uncertain rule like "if P then probably Q ." If in fact the rule "if P then Q " holds, then the generalized compatibility relation $R_1 = \{(P, Q), (\text{not } P, \Theta'')\}$ holds. Let's give it the probability p . If, on the other hand, the rule does not hold, then there is no restriction on $\Theta' \times \Theta''$. This corresponds to $R_2 = \{(\Theta', \Theta'')\}$ which obtains the probability $1 - p$. This defines a conditional belief structure Bel on $\Theta' \times \Theta''$. Let Bel' be a belief function on Θ' and consider a focal set F with basic probability commitment $q(F)$. Then $(F \times \Theta'') \cap C(R_1)$ is a focal set of $Bel' \oplus Bel$. In the sequel, it will be convenient to use the following abbreviation: if $P \subset \Theta'$ and $Q \subset \Theta''$, then $P \cap Q$ or $P \cup Q$ stands for $(P \times \Theta'') \cap (\Theta' \times Q) = P \times Q$ or $(P \times \Theta'') \cup (\Theta' \times Q)$, respectively. Hence, the above focal set may be written as $F \cap ((\text{not } P) \cup Q) = (F \cap (\text{not } P)) \cup (F \cap Q)$. Because we are interested in the belief propagated to Θ'' , we consider the restriction of $Bel' \oplus Bel$ to Θ'' and hence the projection of its focal sets to Θ'' . Now if $F \cap (\text{not } P) \neq \emptyset$, then the focal set above projects to the whole set Θ'' and the focal set F of Bel' induces no particular belief into Θ'' . One may also say that F does not trigger the rule because F does not imply P . If, on the other hand, $F \cap (\text{not } P) = \emptyset$ or $F \subset P$ (i.e., F implies P and thus triggers the rule), then the focal set above projects to $Q \subset \Theta''$ and assigns it the probability $pq(F)$. Other focal sets of Bel' may also contribute to Q . It is then seen that Bel'' is a simple belief function focusing on Q with a probability equaling

$$p \sum_{F \subset P} q(F). \quad (16)$$

As a second, slightly more general example, suppose one wants to model a rule which says that if $P \subset \Theta'$ then a belief on Θ'' is induced which is represented by a belief function Bel with focal sets $Q_j, j = 1, 2, \dots, m$, and basic probability numbers p_j . Such a rule is modeled by generalized compatibility relations $R_j = \{(P, Q_j), (\text{not } P, \Theta'')\}, j = 1, 2, \dots, m$, where each relation holds with prob-

ability p_j . This defines again a belief function Bel on $\Theta' \times \Theta''$ which represents a conditional belief structure. One of the focal sets Q_j , say Q_m , may be equal to Θ'' . Then $R_m = \{(\Theta', \Theta'')\}$. If Bel' is a belief function on Θ' focusing belief 1 to a set $F \subset P$ (i.e., F implies P), then it follows by the same argument as in the first example that the belief function Bel'' which represents the belief transported to Θ'' by the conditional belief structure has exactly the focal sets Q_j and the basic probability numbers p_j as would be expected. If, more generally, Bel' has focal sets F with basic probability numbers $q(F)$, then Bel'' has focal sets Q_j with basic probability numbers (assuming $Q_j \neq \Theta''$)

$$m(Q_j) = p_j \sum_{F \subset P} q(F) \quad (17)$$

and Θ'' with (assuming $Q_m = \Theta''$)

$$m(\Theta'') = p_m \sum_{F \subset P} q(F) + \sum_{F \cap (\text{not } P) \neq \emptyset} q(F). \quad (18)$$

There may be several rules of this type. For $k = 1, 2, \dots, n$, there are rules which say that if $P_k \subset \Theta'$ is true then a belief function on Θ'' with focal sets $Q_{kj}, j = 1, 2, \dots, m_k$ and basic probability numbers p_{kj} holds. Each one of these rules can be modeled according to the second example by a belief function Bel_k on $\Theta' \times \Theta''$. If the rules can be assumed to be *independent*, then

$$Bel = \bigoplus_{k=1}^n Bel_k \quad (19)$$

represents the corresponding conditional belief structure. This case of *independent rules* will be considered in the following subsection.

5.2. Independent Rules

In this subsection, the propagation of belief from Θ' to Θ'' by n independent rules will be discussed in some detail. To simplify, we will consider only simple rules of the type “if P_k then Q_k ” with probability p_k , i.e., P_k implies a simple belief function focusing probability p_k on $Q_k \subset \Theta''$. According to the first example in the previous subsection, this is modeled by a simple belief function Bel_k focusing probability p_k to $(\text{not } P_k) \cup Q_k \subset \Theta' \times \Theta''$. The case of more general rules as in the third example in the previous subsection can be treated along similar lines.

These n rules define a conditional belief structure Bel on $\Theta' \times \Theta''$ according to Eq. (19). Consider a belief function Bel' on Θ' with focal sets F_i and basic probability numbers $q_i, i = 1, 2, \dots, m$. $Bel' \oplus Bel$ then has focal sets

$$\begin{aligned}
& \text{(i)} \quad F_i, i \in I_m, \\
& \text{(ii)} \quad \bigcap_{j \in J} ((\text{not } P_j) \cup Q_j) = \left(\bigcap_{j \in J} (\text{not } P_j) \right) \cup \dots, \\
& \quad \text{for all nonempty subsets } J \subset I_n, \\
& \text{(iii)} \quad F_i \cap \left(\bigcap_{j \in J} (\text{not } P_j) \cup Q_j \right) \\
& \quad = F_i \cap \left(\bigcap_{j \in J} \text{not } P_j \right) \cup \left\{ \bigcup_{H \subset J, H \neq J} F_i \cap (\text{not } P_h) \bigcap_{k \in J-H} Q_k \right\}, \\
& \quad \text{for } i \in I_m, \text{ and for all nonempty subsets } J \subset I_n.
\end{aligned} \tag{20}$$

(Remember that in these formulae, $P \subset \Theta'$ means $P \times \Theta''$ and $Q \subset \Theta''$ stands for $\Theta' \times Q$.)

We are interested in the restriction Bel'' of $Bel' \oplus Bel$ to Θ'' . The focal sets of Bel'' are obtained as projections of the above sets in Eq. (20) to Θ'' . Note that all the sets (i) project to Θ'' . This is also true of the sets (ii) unless

$$\bigcup_{j \in J} P_j = \Theta'. \tag{21}$$

But this means that the rules “if P_j then Q_j ” themselves without any belief on Θ' induce already some belief on subsets of Θ'' . This case will be excluded in the following. Finally, even the sets (iii) project to Θ'' unless

$$F_i \subset \bigcup_{j \in J} P_j. \tag{22}$$

So Eq. (22) is the interesting case. We may say that F_i triggers, in this case, the rules with conditions $P_j, j \in J$. If Eq. (22) holds for no nonempty subset $J \subset I_n$, then F_i triggers no rule and the probability committed to F_i goes to Θ'' .

Let T be the family of subsets $J \subset I_n$ for which Eq. (22) holds and consider a $J \in T$. Because

$$F_i \bigcap_{h \in H} (\text{not } P_h) = \phi \tag{23}$$

for any $H \in T$, it follows that (iii) projects to

$$Pro_J = \bigcup_{H \notin T} \bigcap_{k \in J-H} Q_k \subset \Theta''. \tag{24}$$

Let $T_J = \{K \subset J: J - K \notin T\}$. This is a *monotone* family of sets: if $K \in T_J$ and $K \subset K'$, then also $K' \in T_J$. T_J has one or more *minimal* sets K_1, \dots, K_s , i.e., sets $K_i \in T_J$, but such that no proper subset of K_i belongs to T_J . It is then

clear that it is sufficient to take the union in Eq. (24) only over the minimal sets K_1, \dots, K_S :

$$Pro_J = \bigcup_{h=1}^s \bigcap_{k \in K_h} Q_k. \quad (25)$$

The minimal sets of T_J determine therefore the projection of the focal set (iii). The basic probability number affected by the focal set (iii) and transported to Eq. (25) in the restriction Bel'' is

$$q_i \prod_{j \in J} p_j \prod_{j \in I_n - J} (1 - p_j). \quad (26)$$

This is, of course, not yet the basic probability number of Eq. (25) because other focal sets (iii) may project to the same set and contribute therefore to its probability. It is, furthermore, not excluded that some of the sets of Eq. (25) are empty and that a renormalization is necessary.

This shows how to compute the belief propagated from a belief function Bel' on Θ' to Θ'' by n independent rules. The following three theorems cover three special, but important cases.

THEOREM 5.1: *Suppose that the P_k , $k = 1, 2, \dots, n$, are mutually disjoint and let Bel' be a simple belief function focusing probability q to $F \subset \Theta'$. Let M be the unique smallest subset of I_n such that*

$$F \subset \bigcup_{k \in M} P_k. \quad (27)$$

Then the restriction Bel'' of $Bel \oplus Bel'$ to Θ'' is a simple belief function focusing probability

$$q \prod_{k \in M} p_k \quad (28)$$

to the set

$$\bigcup_{k \in M} Q_k. \quad (29)$$

PROOF: Under the assumptions of this theorem, T consists of M and of all its supersets. For any $J \in T$ and $k \in M$, we have $J - \{k\} \notin T$. The sets $\{k\}$ for $k \in M$ are therefore exactly the minimal sets of T_J . This implies that for any $J \supset M$, Pro_J equals Eq. (29) whereas for all other $J \subset I_n$ we have $Pro_J = \emptyset$. The total probability going to Eq. (29) is then

$$\begin{aligned}
 q \sum_{J \supset M} \prod_{j \in J} p_j \prod_{j \in I_n - J} (1 - p_j) &= q \prod_{k \in M} p_k \sum_{K \subset I_n - M} \prod_{j \in K} p_j \prod_{j \in (I_n - M) - K} (1 - p_j), \\
 &= q \prod_{k \in M} p_k.
 \end{aligned} \tag{30}$$

This proves Eq. (28) and the theorem. ■

It is not difficult to generalize this theorem for a more general belief function than a simple belief function Bel' . The following theorem concerns *consonant* rules:

THEOREM 5.2: Suppose that $P_{k-1} \subset P_k$ and $Q_{k-1} \subset Q_k$ for $k = 2, \dots, n$. Bel' is a simple belief function focusing probability q to $F \subset \Theta'$. Let r be the index such that F is contained in P_r but not in P_{r-1} (if F is subset of P_1 , then $r = 1$; if F is no subset of P_n , then $r = n + 1$). Then Bel'' has the focal sets Q_k for $k = r, r + 1, \dots, n$ (together with Θ'') and they have the basic probability numbers

$$m(Q_k) = qp_k \prod_{h=r}^{k-1} (1 - p_h). \tag{31}$$

Furthermore, for $k = r, r + 1, \dots, n$

$$Bel''(Q_k) = q \left\{ 1 - \prod_{h=r}^k (1 - p_h) \right\}. \tag{32}$$

This theorem says that consonant rules transport simple support into consonant belief.

PROOF: If F is no subset of P_n , then no rule triggers and the whole belief of 1 goes to Θ'' . Otherwise, $r \leq n$ and T contains all $J \subset I_n$ which contain at least one element $j \geq r$. Consider a $J \in T$ and let s be the smallest index greater than or equal to r in J . Then $J - \{j \geq s\} \notin T$ and $K = \{h \in J: h \geq s\}$ belongs to T_J and is in fact the *unique* minimal set of T_J . Thus,

$$Pro_J = \bigcap_{h \in K} Q_k = Q_s. \tag{33}$$

So the $Q_s, s = r, r + 1, \dots, n$, are in fact the focal sets of Bel'' .

The probability affected to $Q_k (k \geq r)$ is

$$m''(Q_k) = q \sum_{J: s=k} \prod_{h \in J} p_h \prod_{h \in I_n - J} (1 - p_h). \tag{34}$$

Define $I(k) = \{1, 2, \dots, r - 1\} \cup \{k + 1, \dots, n\}$. Then it follows from Eq. (34) that

$$\begin{aligned}
 m''(Q_k) &= qp_k \prod_{h=r}^{k-1} (1 - p_h) \sum_{K \subset I(k)} \prod_{h \in K} p_h \prod_{h \in I(k) - K} (1 - p_h), \\
 &= qp_k \prod_{h=r}^{k-1} (1 - p_h).
 \end{aligned} \tag{35}$$

This implies that

$$Bel''(Q_r) = m''(Q_r) = qp_r, \quad (36)$$

and by induction on k ,

$$\begin{aligned} Bel''(Q_k) &= Bel''(Q_{k-1}) + m''(Q_k), \\ &= q \left\{ 1 - \prod_{h=r}^{k-1} (1 - p_h) \right\} + qp_k \prod_{h=r}^{k-1} (1 - p_h), \\ &= q \left\{ 1 - \prod_{h=r}^{k-1} (1 - p_h)(1 - p_k) \right\}, \\ &= q \left\{ 1 - \prod_{h=r}^k (1 - p_h) \right\}. \end{aligned} \quad (37)$$

This proves Theorem 5.2. ■

THEOREM 5.3: Let $P_k, k = 1, 2, \dots, n$, be arbitrary, but suppose Bel' is a simple belief function focusing with probability q to a set $F \subset \Theta'$ which has the property that either $F \subset P_k$ or $F \subset \text{not } P_k$ for any $k = 1, 2, \dots, n$. Let $M = \{k: F \subset P_k\}$ and let Bel''_k be simple belief functions focusing probability p_k on $Q_k \subset \Theta''$. Then

$$Bel'' = q \bigoplus_{k \in M} Bel''_k. \quad (38)$$

M represents here the set of rules which are triggered with probability q . The theorem says then essentially that the conditional belief transported to Θ'' by the triggered rules can be combined on Θ'' by Dempster's rule.

PROOF: T consists of all $J \subset I_n$ which contain at least one element of M . Fix a $J \in T$. Then $K \in T_J$ if and only if $J-K$ does not intersect with M . There is therefore a unique minimal set $K = J \cap M$ in T_J . This implies that

$$Pro_J = \bigcap_{k \in J \cap M} Q_k. \quad (39)$$

Equation (39) means that the focal sets of Bel'' (other than Θ'') are all formed by intersections

$$\bigcap_{h \in H} Q_h \quad (40)$$

over nonempty subsets H of M . Some of the intersections (40) may be empty. But the probability affected by Eq. (40) for a given H is obtained as the sum over the probabilities affected to the sets (iii) in Eq. (20) above for which $J \cap M = H$:

$$\begin{aligned}
& \sum_{J: J \cap M = H} q \prod_{j \in J} p_j \prod_{j \in I_n - J} (1 - p_j) \\
&= q \prod_{h \in H} p_h \prod_{h \in M - H} (1 - p_h) \sum_{K \subset I_n - M} \prod_{k \in K} p_k \prod_{k \in (I_n - M) - K} (1 - p_k), \quad (41) \\
&= q \prod_{h \in H} p_h \prod_{h \in M - H} (1 - p_h).
\end{aligned}$$

These probabilities have to be conditioned (renormalized) using the empty sets of Eq. (40), and they have also to be regrouped for identical sets of Eq. (40) arising from different H . But that is (up to the factor q) exactly as the focal sets and basic probability numbers of the orthogonal sum in Eq. (38) would be computed. This proves the theorem. ■

6. CONCLUSION

It has been shown in this paper that uncertain “if . . . then . . .” rules can be modeled within Shafer’s mathematical theory of evidence by beliefs about compatibility relations between two or more variables or frames of discernment. More generally, conditional belief structures representing conditional beliefs like “if some proposition in Θ' holds, then there is a certain belief function on Θ ” can be represented by belief functions on the product set $\Theta' \times \Theta$ which represent also belief about compatibility relations. The evidence contained in such conditional belief functions can then be combined with beliefs about variables (represented by belief functions) using the usual Dempster rule.

The concepts developed here fit well into the framework developed in the paper of Shafer et al. [9]. Whereas these authors stress the point of view of partitions (which is preferable for basic, theoretical considerations), in this paper the multivariate approach is used (which is necessary for practical modeling). Shafer et al. [9] show that propagating belief in a system of conditional belief structures can be computationally simplified if a certain topological property is satisfied. This is so because in these cases the overall computation can be broken up into many local computations. For these local computations, results like those obtained in Subsection 5.2 may be useful.

If that particular topological property is absent, then the computations for propagating belief become more complicated. In fact very little is known whether there are other cases where the computations can be simplified (see, however, Kong [3]). As Theorem 4.1 in Subsection 4.3 shows, there are some relations to *network reliability problems*. In the latter domain, there are some results on how network topology can be exploited in some cases to get efficient computational procedures. Perhaps some of these results could be used to construct efficient procedures for propagating belief in non-Markovian networks.

References

1. Dempster, A.P. (1967). Upper and lower probabilities induced by a multivalued mapping. *Annals of Mathematical Statistics* 38: 325–339.
2. Garvey, T.D. (1986). Evidential reasoning for land-use classification. A.I. Center, SRI Int., Menlo Park, CA 94025.
3. Kong A. (1986). *Multivariate belief functions and graphical models*. Doctoral dissertation, Department of Statistics, Harvard University, Cambridge.
4. Lowrance, J.D. (1986). Automating argument construction. A.I. Center, SRI Int., Menlo Park, CA 94025. *Proceeding Workshop on Assessing Uncertainty*, Naval Postgraduate School, Monterey, Nov. 1986.
5. Lowrance, J.D. & Garvey, T.D. (1983). Evidential reasoning: an implementation for multisensor integration. A.I. Center, SRI Int., Menlo Park, CA 94025.
6. Lu, S.Y. & Stephanou, H.E. (1984). A set theoretic framework for the processing of uncertain knowledge. *Proceedings of the National Conference on Artificial Intelligence*, University of Texas at Austin, TX, p. 216–221.
7. Pearl, J. (1986). Fusion, propagation, and structuring in belief networks. *A.I.* 28: 241–288.
8. Shafer, G. (1976). *A mathematical theory of evidence*. New Jersey: Princeton University Press.
9. Shafer, G., Shenoy, P.P., & Mellouli, K. (1986). Propagating belief functions in qualitative Markov chains. School of Business Working Paper No. 186. The University of Kansas, Lawrence.
10. Shafer, G. & Tversky A. (1985). Languages and designs for probability judgment. *Cognitive Science* 9: 309–339.